The Monopolistic Competition Revolution in Retrospect

Steven Brakman Ben J. Heijdra

December 2002

Abstract

In this paper we argue that there have been two monopolistic competition revolutions. The first was started by Joan Robinson and Edward Chamberlin in the 1930s but failed to have much impact on economic theory. The second was initiated by Avinash Dixit and Joseph Stiglitz in the early 1970s. Their revolution succeeded because it yielded an analytically tractable model of Chamberlinian monopolistic competition (the "large group" case). This model has been used in such diverse fields as international trade theory, economic geography, economic growth theory, and macroeconomics. Its popularity shows no sign of decline.

This paper forms the introduction to the forthcoming book entitled The Monopolistic Competition Revolution in Retrospect edited by S. Brakman and B.J. Heijdra (Cambridge University Press, 2003). This book is based on the papers presented at the SOM conference which was held at the University of Groningen, on October 30-31, 2000.

1 Introduction

> In speaking of theories of monopolistic or imperfect competition as "revolutions," I know in advance that I shall provoke dissent. There are minds that by temperament will define away every proposed revolution. For them it is enough to point out that Keynes in 1936 had some partial anticipator in 1836. Newton is just a guy getting too much credit for the accretion of knowledge that covered centuries. A mountain is just a high hill; a hill merely a bulging plain. Such people remind me of the grammar-school teacher we all had, who would never give 100 to a paper on the ground that "No one is perfect." Samuelson (1967, p. 138).

Edward Hastings Chamberlin is the author of one of the most influential works of all time in economic theory—*The Theory of Monopolistic Competition*, which entered its eighth edition in 1962. Along with Lord Keynes's *General Theory*, it wrought one of the two veritable revolutions in economic theory in this century (Dust cover text of Kuenne (1967)).

Although we stress the importance of the contribution by Avinash Dixit and Joseph Stiglitz (1977) throughout this book, the history of monopolistic competition is much longer than the past 25 years or so and goes back at least 70 years. The success of the Dixit-Stiglitz model of monopolistic competition might have come as a surprise to students of the history of economic thought, as it was by no means the first attempt to deal with imperfect markets or monopolistic competition. However, where the earlier attempts failed the Dixit-Stiglitz approach turned out to be very successful and has the potential "for classic status" (see Neary[1], this volume).

In this introduction we will briefly review the two waves of literature on monopolistic competition theory, namely the one that started in 1933 and the one that commenced in 1977. The claim of this book is that the second attempt to model monopolistic competition was far more successful than the first one, essentially because the second attempt introduced a formalization that had all the relevant characteristics of monopolistic competition but was still relatively easy to handle.

This collections of papers will show that the re-formulation by Dixit and Stiglitz has contributed significantly to many areas of research; the main ones being international trade theory, macroeconomics, growth theory and economic geography. But even today the concept of monopolistic competition is not always appreciated. As David Kreps puts it in his influential micro textbook "...were it not for the presence of this theory in most lower level texts we would ignore it here altogether" (1990, p. 344). Kreps dismisses monopolistic competition as being too unrealistic, and challenges his readers to come up with at least one sector that could convincingly be described by monopolistic competition. This collection of essays, however, takes for granted that the Dixit-Stiglitz reformulation of monopolistic competition has become very successful, and asks why that is the case. This does not mean that the authors of the essays are uncritical about the model. The aim of this collection of essays is to show why the model has become mainstream in such a short period of time and what we can expect from future developments regarding the modelling of imperfect markets.

This introductory essay is organized as follows. In Section 2 we briefly discuss the literature predating the first monopolistic competition revolution. This literature strongly hinted at the importance of increasing returns to scale and imperfect market forms but was unable to come up with a satisfactory model in which both phenomena could play a meaningful role.

In Section 3 we briefly discuss (what we call) the first monopolistic competition revolution, namely the one that was started by Edward Hastings Chamberlin and Joan Robinson in the 1930s. We show that by the mid 1960s most (but not all) leading economists had come to the conclusion that the Chamberlin-Robinson revolution had essentially failed. In our view, there are two reasons for this lack of acceptance of the theory. First, the *timing* of the first revolution was unfortunate in that it coincided with the Great Depression and the emergence

[1] According to Peter Neary, "the first step on the road to classic status [is]: to be widely cited but never read. (The second step, to be widely quoted but never cited)."

of the Keynesian revolution in macroeconomics. Second, and perhaps more importantly, Chamberlin and co-workers failed to come up with a canonical model embodying the key elements of the theory. It was not so much Chamberlin's ideas that were rejected but rather his *modelling approach* that was deemed to be unworkable.

In Section 4 we turn to the second monopolistic competition revolution, namely the successful one that was started in the mid 1970s by Dixit, Stiglitz, and Michael Spence. The timing of this second revolution was much better—there were no major theory revolutions or economic catastrophes going on at the time. More importantly, the second revolution caught on because Dixit and Stiglitz managed to come up with a canonical model of monopolistic competition. We present a very simple version of the Dixit-Stiglitz model and show how it manages to capture the key Chamberlinian insights.

Finally, in Section 5 we present a broad overview of the papers in this book.

2 Precursory Thoughts on Imperfect Competition[2]

By the end of the nineteenth century two market forms dominated the discussion of economic analysis of the time, namely monopoly and perfect competition. The former assumes a single firm with exclusive control over its output and the market, resulting in profits that are larger than in any other market form. In contrast, the latter assumes a large number of sellers of a homogeneous product, where each individual firm has no control over its price. Free entry and exit of firms ensures that long-run profits are zero. Perfect competition was introduced to show that in some sense it is optimal and in fact represents an end-state, meaning that competition between buyers or sellers has come to an end and neither party can increase utility or profits by changing its behaviour. Changes only occur if exogenous variables change, but the question then becomes how fast and under what circumstances the new equilibrium will be reached. Competition might not actually lead to the blissful state but market forces are always pointing the economy in the right direction.[3] Monopoly by contrast maximizes profits of the firm but from a social point of view is sub-optimal.

This state of affairs is reflected in Alfred Marshall's *Principles of Economics*, that presented these two market forms as the basic analytical tools to analyze markets. Other market forms are hybrids in between these two polar cases.[4] Mainstream economics did not bother

[2]Our historical overview is rather succinct due to space considerations. Interested readers are referred to Triffin (1940), Eaton and Lipsey (1989, pp. 761-766) and Archibald (1987, pp. 531-534) for more extensive surveys.

[3]As Arrow and Debreu showed, in general the conditions for a unique and stable (Walrasian) equilibrium are that (1) production is subject to constant or diminishing returns to scale, (2) commodities are substitutes (meaning that a price increase raises the demand for other products), (3) external effects are absent, and (4) there is a complete forward market for all goods. Especially assumptions (1) and (3) are dropped in monopolistic competition.

[4]However, according to Schumpeter, Marshall "had no theory of monopolistic competition. But he pointed toward it by considering a firm's Special Market." (Schumpeter, 1954, p. 840).

too much to analyze imperfect market forms, because "the large majority of cases that occur in practice are nothing but mixtures and hybrids of these two" (Schumpeter, 1954, p. 975).

However, Marshall was aware that other market forms were not simple combinations of perfect competition and monopoly. The special nature of imperfect markets were conveyed to him in the form of the duopoly models developed by Cournot, Bertrand and Edgeworth in the second half of the nineteenth century. Especially the analysis of Cournot (1838) was important for him, as it handed to him the apparatus to analyze market forms in the first place. The problem with these models was that the results depended very much on special assumptions, and this might have precluded Marshall from contributing to this line of theory himself. Although Marshall did not develop his own theory of imperfect competition, his awareness of the so-called *Special Markets* paved the way for later theories of imperfect competition developed by Chamberlin and Robinson.

Notwithstanding some lip-service to the theory of imperfect competition, perfect competition dominated the analysis during this time and other market forms were considered to be 'imperfect.' However, in perfect competition, where each seller or buyer has no influence on market prices, there is no longer room for individual competition, and forces leading to *industry* growth are absent. The difficulty was then to reconcile the theory of the market and that of the individual firm. Simple observation of reality often contradicted the conclusions of (partial) supply and demand analysis: diminishing returns for the individual firm is not an obstacle to expand production. And, more often than not, average costs are diminishing at the point were firms stop expanding output. This state of affairs troubled Marshall, as decreasing (average) cost curves are incompatible with perfect competition. Marshall tried to solve this by introducing diminishing returns for the individual firm (for individual firms, production factors are in fixed supply), and external economies for the whole industry. The introduction of external economies of scale at the industry level ensured that the competitive equilibrium could be rescued. The central idea is that external economies of scale create an interdependence between supply curves; the combined supply of all firms reduces industry costs and ensures that the combination of lower prices and increased supply can be an equilibrium. External economies of scale are compatible with an industry equilibrium, because an increase in demand will still increase the price for *individual* firms, as the marginal cost curve of each firm is upward sloping and each firm is operating at the minimum of its average cost curve. The price increase could stimulate new firms to enter the market, reducing (average) costs and raising combined supply. With internal economies of scale a market equilibrium is not possible as each individual firm can always undercut its rivals.

According to Marshall whether or not external economies could be encountered in practice depended on the general characteristics of an industry and the environment of the industry, like the localization of an industry. In Marshall's words:

> ...subsidiary trades grow up in the neighbourhood, supplying it with implements

> and materials, organizing its traffic, and in many ways conducing to the economy of its material...the economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind,...subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbours, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expenses... (Marshall, 1920, p. 225)

In modern jargon the linkages described in this quotation are so-called backward and forward linkages; the backward linkage is that firms use other firms' output as intermediate production factors, the forward linkage is that its own product is also used as an intermediate production factor by others.

Furthermore, according to Marshall a thick labour market also benefits firms:

> Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market. (Marshall, 1920, p. 225-6)

These factors combined explain industry growth and show why:

> ...the mysteries of the trade become no mysteries; but are as it were in the air...if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas (Marshall, 1920, p. 225)

For Marshall, however, his analysis of external economies created an additional problem, because he thought that internal economies of scale were at least as important as external economies. In the presence of internal economies of scale the growth of an industry would benefit the largest firms (and create monopolies) and thus change the competitive forces within such an industry. For consistency reasons Marshall had to introduce the concept of the representative firm to deal with this incompatibility. By introducing the representative firm, perfect competition and (external) economies of scale could be made consistent. This was considered necessary since the economic profession at the time considered perfect competition the rule and monopolies the exception. But again in this case, as with perfect competition, strategic interaction between firms has been assumed away because firms are by assumption "representative" for the whole industry.

But the consistency problems in Marshall's analysis of the market were not solved even by the representative firm. Marshall's famous period analysis assumed that in the long run the supply curve was a straight line. And this means that *in the long run* the volume of production of an individual firm is indeterminate: there is no unique intersection of the

supply curve and a given price. So, Marshall's theory of perfect competition has no way of dealing with situations where the (long-run) marginal costs are constant (or declining in the presence of economies of scale). This state of affairs was most poignantly put forward by Sraffa (1926). According to Sraffa market imperfections due to returns to scale are not simple frictions, "but are themselves active forces which produce permanent and even cumulative effects." And he added yet another problem. Declining marginal costs would imply that the market is served by a single firm. But, according to Sraffa, in practice firms operate under declining marginal costs without monopolizing the whole market. According to him, the combination of a declining supply curve and a negatively sloped demand curve limits the size of production. The idea behind a declining demand curve is that buyers are not indifferent between different suppliers. Each firm has his own *special market*; products are usually imperfect substitutes and have their own *special* characteristics.

In a sense Sraffa added to the confusion rather than solving the problem of combining increasing returns and the theory of market competition. The error Sraffa made was that he did not distinguish between price and marginal revenue, which was remarkable because the concept of marginal revenue had already been developed in a mathematical appendix in Marshall's *Principles*, in which he restates the monopoly theory developed by Cournot.[5] This was pointed out (again) by Harrod in 1930.[6] For Marshall it was a minor issue and he did not make use of this instrument any further, because he did not need it in his analysis of perfect competition.

This was broadly speaking the state of affairs in the 1920s and 1930s. It was realized that the existence of economies of scale (of one sort or another) implies imperfect market forms, but it remained difficult to construct a satisfactory equilibrium concept for such imperfect market forms. On the one hand there was perfect competition, and on the other hand there was monopoly. Other market forms were considered to be some kind of hybrid of these two extreme forms of competition. So, one could suffice to analyze the two extreme cases in treating all other forms as an implicit mix of the two fundamental forms of competition. But no satisfactory theory of the market existed in which constant or declining marginal and average costs could be made consistent with market equilibrium. This led in the 1930s to a new theory of price determination. One can agree with Schumpeter (1954, p. 1150) that the confusion caused by Marshall was a very fertile one.[7] Marshall's analysis of the firm and

[5] Marshall casts his analysis in terms of net revenue, and only implicitly discusses marginal revenue. The concept of marginal revenue had to be re-invented (Robinson, 1933). This is even more surprising considering that Cournot already used the concept of marginal revenue in 1838, and derived the familiar first-order condition for profit maximization: marginal revenue equals marginal cost (Cournot, 1838).

[6] See Harrod (1967) for a review of his thoughts on this matter.

[7] Chamberlin, for example, attributed the origins and inspiration of his theory to the famous Taussig-Pigou controversy on railway rates which took place around 1900. This controversy was about the explanation of different railway rates. Taussig tried to fit different railway rates into the Marshallian theory of (competitive) joint supply by assuming that a unit rail supply is not homogenous and that different demand elasticities for different stretches of railway result in different prices. In contrast, Pigou stated that it was not an issue of

economies of scale led him to develop the concept of the representative firm which invited a lively discussion on market equilibrium and returns to scale and this set the stage for the analysis of monopolistic competition.

3 Monopolistic Competition in the 1930s

In 1933 two books appeared that changed the way economists dealt with imperfect competition, namely Joan Robinson's *The Economics of Imperfect Competition* and Edward Hastings Chamberlin's *The Theory of Monopolistic Competition.* Although Robinson revived the marginal revolution, in general Chamberlin is considered to be "the true revolutionary" (Blaug, 1997, p. 376).[8] This radical new analysis was a first answer to the question that was raised in 1926 by Sraffa: is it possible in a market characterized by monopolistic competition and declining average and marginal costs to reach an equilibrium. The figure below illustrates the equilibrium in the monopolistic equilibrium. Chamberlin makes four basic assumptions (Bishop, 1967, p. 252):

- The number of sellers in a group of firms is sufficiently large so that each firm takes the behaviour of other firms in the group as given (Cournot-Nash assumption).
- The group is well defined and small relative to the economy.
- Products are physically similar but economically differentiated: buyers have preferences for all types of products.
- There is free entry and exit.

The monopolistic elements are all those elements that distinguish a product from another product and give the firm some market power; "...each 'product' is rendered unique by the individuality of the establishment in which it is sold, including its location (as well as by trade marks, qualitative differences, etc); this is its monopolistic aspect" (Chamberlin, 1933, p. 63). The large number of firms in the market and the possibility of entry and exit of many firms provide the competitive elements; "Each (product) is subject to the competition of other 'products' sold under different circumstances and at other locations; this is its competitive aspect" (1933, p. 63).

We illustrate the Chamberlinian model with the aid of Figure 1.[9] We assume that all actual and potential suppliers in the group face the same demand and cost conditions and depict the situation for one particular firm in isolation. There are two demand curves in

heterogeneity, but of monopoly coupled with the conditions necessary for price discrimination which could explain price differences. In general it is thought that Pigou won the debate.

[8] Moreover, the history of Chamberlin's seminal work dates back to 1921—see the remarks by Schumpeter (1954, p. 1150).

[9] This diagram is adjusted from Bishop (1967, p. 252).

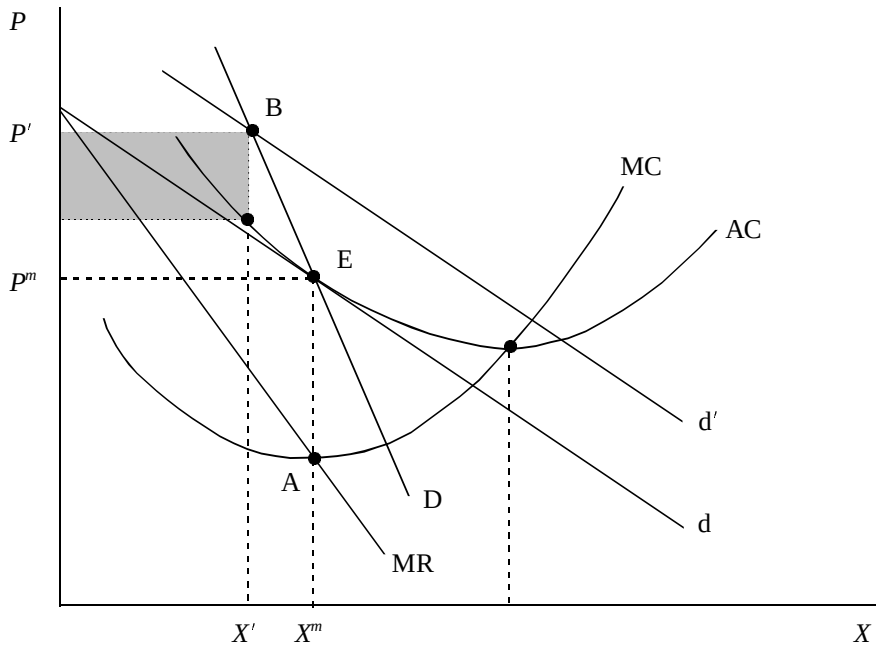

Figure 1: Chamberlinian monopolistic competition equilibrium

the diagram. The individual firm under consideration faces demand curve d. This curve represents the firm's price-sales combinations under the assumption that all other firms in the group keep their prices unchanged. Archibald calls this the 'perceived' demand curve (1987, p. 532). The steeper curve labelled D is the demand facing each firm if all firms in the group set their prices identically. Archibald (1987, p. 532) refers to this curve as the 'share-of-the-market' demand curve. As usual MR is marginal revenue (associated with the perceived demand curve d), AC is the firm's average cost, MC is marginal cost, P is the price of the differentiated commodity, and X is the volume of sales.

The Chamberlinian equilibrium under free entry/exit of firms is at point E, where the price is P^m and output is X^m. Point E is the equilibrium because (a) the individual firm attains an optimum in that point, and (b) there are no unexploited profit opportunities, excess profits are exactly zero, and no entry/exit of firms takes place. The validity of these requirements can be demonstrated as follows. The individual firm maximizes its profit, taking as given the demand curve d. It finds the optimum point by equating marginal revenue and marginal cost (see point A directly below point E). In point E the demand curve d is tangent to the average cost curve, AC, so the firm makes zero profits. This is the famous Chamberlinian tangency condition. Since all firms are identical, no firm makes profits or losses and there is no entry or exit of firms.

Chamberlin (1933, p. 91) also sketched the adjustment process towards the equilibrium point. Assume that all firms in the group are initially operating along the demand curve d′ at point B, set a price of P', and produce a quantity X'. At this price-output combination, each firm would make a positive profit equal to the shaded area in Figure 1. But point B cannot be an equilibrium. Indeed, in that point the individual firm will have an incentive to lower its price (and increase its profits) by moving to the right along the d′ curve (recall that each firm operates under the assumption that its competitors will continue to charge P'). But each firm has exactly the same incentives, so they will all follow suit and cut their prices. As a result the d′ curve will shift down along the D curve towards the Chamberlinian equilibrium at point E.[10]

Obviously, due to the downward sloping individual demand curve, there is a difference between equilibrium average cost and minimum average cost in the Chamberlinian equilibrium. This implies that there are unexploited economies of scale and the question arises whether this represents a waste of resources. The answer to this question is both yes and no. Yes, in the sense that indeed there is excess capacity and no, in the sense that product differentiation introduces variety and this expands the extent of consumer choices and thereby welfare. As Eaton and Lipsey put it, "in a society that values diversity, there is a trade-off between economizing on resources, by reducing the costs of producing existing products, and satisfying the desire for diversity, by increasing the number of products" (1989, p. 763). We will return to this topic in more detail when discussing the second monopolistic competition revolution.

Given the elegance of the monopolistic competition model it is surprising to see how little influence it had on economic theory. The first attacks on the early monopolistic competition revolution came from Hicks (1939, pp. 83-85) and somewhat later from Stigler (1949) and Friedman (1953). Hicks rejected the theory because he was unable to translate it into a workable model. Stigler (1949) rejected the theory for methodological reasons. He claimed that the predictions derived from the theory of monopolistic competition are not very different from those of perfect competition. Occam's razor then suggests that perfect competition should be favoured over monopolistic competition, a line of reasoning to which Friedman also adheres. It was put forward even more strongly by Archibald (1961, p.14): "The theory is not totally empty, but very nearly so" (see also Samuelson (1967) for a further discussion of this debate). In addition Stigler raised an important point by noting that:

> Professor Chamberlin's failure to construct an analytical system capable of dealing informatively with his picture of reality is not hard to explain. The fundamental fact is that, although Chamberlin could throw off the shackles of Marshall's view of economic life, he could not throw off the shackles of Marshall's view of economic

[10] Note that the position of the D curve depends on the number of firms in the group. In Figure 1, D is consistent with the Chamberlinian equilibrium at E. As a result, the thought experiment conducted above does not prompt entry of firms. It just shows that E is the only conceivable Chamberlinian (Cournot-Nash) equilibrium.

> analysis. Marshall's technique was appropriate to the problem set to it: it deals informatively and with tolerable logic with the world of competitive industries and monopolies. But it is lost in the sea of diversity and unsystematism, and Chamberlin is lost with it. Stigler (1949, p. 22).

Archibald (1987, p. 532) mentions two further criticisms that were raised against the Chamberlinian model. First, the notion of the 'group' (of products) was ill-defined. In the common definition, goods belong to a group if (1) the cross-elasticity of demand beween these goods is 'high' and (2) the cross-elasticity between goods in the group and all other goods is 'low'. The problem with this definition is that there is no logical way to determine what is a high elasticity and what is a low one. Second, Kaldor (1934, 1935) suggested very early on that reality may be better approximated by a market structure with chains of overlapping oligopolies (localized competition) than by Chamberlin's monopolistically competitive structure. Of course, in such an oligopolistic setting the Cournot-Nash assumption is clearly untenable.

Not surprisingly, in well-known textbooks that appeared in the 1960s and 1970s, monopolistic competition is only briefly mentioned, if at all—see for example Henderson and Quandt (1971) and Malinvaud (1972). Akerlof (2002, p. 413) recollects about this period that, "...monopolistic competition and Joan Robinson's equivalent were taught in graduate and even undergraduate courses. However, such 'specific' models ...were presented not as central sights, but instead as excursions into the countryside, for the adventurous or those with an extra day to spare."

The *Festschrift* that was published in honour of Chamberlin also paints a rather bleak picture. Harry Johnson, for example, not only observes that the theory had by 1967 no discernible impact on the theory of international trade, but continues that "...some beginnings have been made towards the analytical and empirical application of monopolistic competition concepts; but the work has been very much ad hoc, and much synthesizing remains to be done" (1967, p. 218). What is needed is an "operationally relevant analytical tool capable of facilitating the quantification of those aspects of real-life competition..." (1967, p. 218).

But not only Johnson is rather sceptical on the contribution of monopolistic competition; other contributors seem to have to same opinion. Fellner, for instance, concludes that these models are convenient tools of exposition "on specific symmetry assumptions...In situations lacking these traits of symmetry...[they] lose much of their usefulness" (1967, p. 29) and Tinbergen (1967) observes that the influence on econometrics and macroeconomics is limited.

Only Paul Samuelson is more positive, though still on the defensive, as the following rather lengthy quotation shows:

> "If the real world displays the variety of behavior that the Chamberlin-Robinson models permit—and I believe the Chicago writers are simply wrong in denying that these important empirical deviations exist—then reality will falsify *many* of

> the important qualitative and quantitative *predictions* of the competitive model. Hence, by the pragmatic test of prediction adequacy, the perfect-competition model fails to be an adequate approximation...The fact that the Chamberlin-Robinson model is 'empty' in the sense of ruling out few empirical configurations and being capable of providing only formalistic descriptions, is not the slightest reason for abandoning it in favor of a 'full' model of the competitive type *if reality is similarly* 'empty' and 'non-full." (1967, p. 108n)

Samuelson concludes that "Chamberlin, Sraffa, Robinson, and their contemporaries have led economists into a new land from which their critics will never evict us" (1967, p. 138).

It might have come as a surprise, even to a relative optimist like Paul Samuelson, that the theory of monopolistic competition was given a new lease on life so quickly. Indeed, less than a decade after the 1967 Chamberlin festivities, Dixit and Stiglitz (1977) managed to again place monopolistic competition theory on the centre stage.

4 The Second Monopolistic Revolution

As we pointed out in the previous section, the monopolistic competition revolution by no means started with the seminal article by Dixit and Stiglitz (1977), but had already had a long (and somewhat troublesome) history. However, one of the reasons why we have gathered the collection of papers in the present volume is that we claim that the second monopolistic competition revolution has been much more successful than the first. The reason for this success is that Dixit and Stiglitz managed to formulate a canonical model of Chamberlinian monopolistic competition which is both easy to use and captures the key aspects of Chamberlin's model. Though it is by now widely recognized that the Dixit-Stiglitz approach is somewhat unrealistic, it has nevertheless become the "workhorse model" incorporating monopolistic competition, increasing returns to scale, and endogenous product variety. As is stressed by Peter Neary in this volume, the main contributions of the Dixit-Stiglitz model are:[11]

- The definition of an industry (or large group of firms) is simplified: all product varieties are symmetric and are combined in a constant-elasticity-of-substitution (CES) aggregation function (see below).
- Overall utility is separable and homothetic[12] in its arguments, implying that we can use

[11] There are actually two models in the original Dixit-Stiglitz (1977) paper, which they label, respectively, the Constant Elasticity Case and the Variable Elasticity Case. The first model has become known as *the* Dixit-Stiglitz model. Note that both models have been used in international trade theory, notably Krugman (1979, 1980).

[12] This is the main distinction from the model developed by Spence (1976), who uses a quasi-linear utility specification.

a two-stage budgeting procedure. In the first stage usually a Cobb-Douglas specification is used, and in the second stage a CES-utility function is applied.

- On the production side, technology features increasing returns to scale at firm level. The typical formulation models the average cost curve as a rectangular hyperbola. All firms are symmetrical.

In the remainder of this section we present a very simple version of the Dixit-Stiglitz model and characterize its key properties. Readers who are familiar with the model may skip this section and proceed directly to section 5 below.

4.1 The model

4.1.1 Preferences

There are two sectors in the economy. The first sector produces a homogeneous good under constant returns to scale and features perfect competition. The second sector consists of a large group of monopolistically competitive firms who produce under increasing returns to scale at firm level. The utility function of the representative household[13] is Cobb-Douglas:

$$U = Z^{\delta} Y^{1-\delta}, \qquad 0 < \delta < 1, \tag{1}$$

where U is utility, Z is consumption of the homogeneous good, and Y is the consumption of a *composite* differentiated good. This composite good consists of a bundle of closely related product "varieties" which are close but imperfect substitutes for each other. Following the crucial insights of Spence (1976) and Dixit and Stiglitz (1977), a convenient formulation is as follows:

$$Y \equiv \left[\sum_{i=1}^{N} X_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \qquad 1 < \sigma \ll \infty, \tag{2}$$

where N is the existing number of different varieties, X_i is consumption of variety i, and σ is the Allen-Uzawa cross-partial elasticity of substitution. Intuitively, the higher is σ, the better substitutes the varieties are for each other.[14] In this formulation, $1/(\sigma - 1)$ captures the notion of "preference for diversity" (or "love of variety") according to which households prefer to spread a certain amount of production over N differentiated goods rather than concentrating it on a single variety (see Bénassy (1996) for this definition).[15]

[13] There is a large number of identical households. To avoid cluttering the notation, however, we normalize the number of households to unity.

[14] In the limiting case, as σ approaches infinity, the varieties are *perfect* substitutes, i.e. they are identical goods from the perspective of the representative household.

[15] In formal terms *average* PFD can be computed by comparing the value of composite consumption (Y) obtained if N varieties and X/N units per variety are chosen with the value of Y if X units of a single variety

The household faces the following budget constraint:

$$\sum_{i=1}^{N} P_i X_i + P_Z Z = I, \tag{3}$$

where P_i is the price of variety i, P_Z is the price of the homogeneous good, and I is household income (see below).

The household chooses Z and X_i (for $i = 1, \cdots, N$) in order to maximize utility (1), subject to the definition of composite consumption (2) and the budget constraint (3), and taking as given the goods prices and its income. By using the convenient trick of two-stage budgeting we obtain the following solutions:[16]

$$P_Z Z = \delta I, \tag{4}$$

$$P_Y Y = (1-\delta) I, \tag{5}$$

$$X_i = (1-\delta) \left(\frac{P_i}{P_Y}\right)^{-\sigma} \left(\frac{I}{P_Y}\right), \qquad (i = 1, \cdots, N), \tag{6}$$

where P_Y is the *true price index* of the composite consumption good Y:

$$P_Y \equiv \left[\sum_{i=1}^{N} P_i^{1-\sigma}\right]^{1/(1-\sigma)}. \tag{7}$$

Intuitively, P_Y represents the price of one unit of Y given that the quantities of all varieties are chosen in an optimal (utility-maximizing) fashion by the household.[17] Equations (4)-(5) feature the usual result that income spending shares on Z and Y are constant for the Cobb-Douglas utility function. Equation (6) is the demand curve facing the producer of variety i. It features a constant price elasticity, i.e. $-\frac{\partial X_i}{\partial P_i}\frac{P_i}{X_i} = \sigma$.[18]

---

are chosen ($N = 1$):

$$\text{average PFD} \equiv \frac{Y(X/N, X/N, ..., X/N)}{Y(X, 0, ..., 0)} = N^{1/(\sigma-1)}. \tag{a}$$

The elasticity of this function with respect to the number of varieties represents the *marginal* taste for additional variety which plays an important role in the monopolistic competition model. By using (a) we obtain the expression for the marginal preference for diversity (MPFD):

$$\text{MPFD} = \frac{1}{\sigma - 1}. \tag{b}$$

[16] For a pedestrian derivation of such expressions, see for example Brakman, Garretsen, and van Marrewijk (2001, ch. 3).

[17] Formally, P_Y is defined as follows:

$$P_Y \equiv \left\{ \min \sum_{i=1}^{N} P_i X_i \text{ subject to } \left[\sum_{i=1}^{N} X_i^{(\sigma-1)/\sigma}\right]^{\sigma/(\sigma-1)} = 1 \right\}.$$

[18] In deriving this elasticity, we follow Dixit and Stiglitz (1977) by ignoring the effect of P_i on the price index P_Y. See Yang and Heijdra (1993), Dixit and Stiglitz (1993), and d'Aspremont et al. (1996) for a further discussion of this point.

Note that equation (6) provides a formal definition for the individual firm's perceived demand curve (i.e. the d curve in Figure 1). To derive the industry demand curve (the D curve) we postulate symmetry (see below), set $P_i = P$ and $X_i = X$ (for all $i = 1, \cdots, N$), and obtain from (6):

$$X = \frac{1}{N}(1-\delta)\frac{I}{P}. \tag{8}$$

Whereas the d curve features a price elasticity of σ (which exceeds unity by assumption), the Cobb-Douglas specification ensures that the D curve is unit elastic, i.e. the *industry* demand curve is less elastic than the demand curve facing individual firms, as was asserted by Chamberlin (1933), and is illustrated in Figure 1 where the D curve intersects the d curve from above. .

4.1.2 Technology and Pricing

The supply side of the economy is as follows. There is one factor of production, labour, which is perfectly mobile across sectors and across firms in the monopolistic sector. As a result, there is a single wage rate which we denote by W. Production in the homogeneous goods sector features constant returns to scale and technology is given by:

$$Z = \frac{L_Z}{k_Z}, \tag{9}$$

where L_Z is the amount of labour used in the Z-sector and k_Z is the (exogenous) technology index in that sector. The Z-sector operates under perfect competition and marginal cost pricing ensures that there are zero profits and the price is set according to:

$$P_Z = k_Z W. \tag{10}$$

Production in the monopolistically competitive Y-sector is characterized by internal economies of scale. Each individual firm i uses labour to produce its product variety and faces the following technology:

$$X_i = \begin{cases} 0 & \text{if } L_i \leq F \\ (1/k_Y)[L_i - F] & \text{if } L_i \geq F \end{cases}, \tag{11}$$

where X_i is the marketable output of firm i, L_i is labour used by the firm, F is fixed cost in terms of units of labour, and k_Y is the (constant) marginal labour requirement. The formulation captures the notion that the firm must expend a minimum amount of labour ("overhead labour") before it can produce any output at all (see Mankiw, 1988, p. 9). As a result, there are increasing returns to scale at firm level as average cost declines with output.[19]

[19] Note that (11) implies that the average cost curve of active firms is a hyperbola. This is standard in the Dixit-Stiglitz model. Most graphical presentations of the Chamberlinian model use U-shaped average cost

The profit of firm i is denoted by Π_i and equals revenue minus total (labour) costs:

$$\Pi_i \equiv P_i X_i - W\left[k_Y X_i + F\right]. \tag{12}$$

The firm chooses its output in order to maximize profit (12) subject to its price-elastic demand curve (6), ignoring the effects its decisions may have on P_Y and/or I (see footnote 18). The first-order condition for this optimization problem yields the familiar markup pricing rule:

$$P_i = \mu W k_Y, \qquad \mu \equiv \frac{\sigma}{\sigma - 1}, \tag{13}$$

where μ (> 1) is the gross markup of price over marginal cost.

4.1.3 Chamberlinian Equilibrium

The key thing to note is that the model is completely symmetric. According to (13), all active firms face the same price elasticity (and thus adopt the same markup), pay the same wage rate, and face the same technology. Hence, all firms set the same price, i.e. $P_i = P$ for all i. But this means, by (6) and (11)-(12), that output, labour demand, and the level of profit are also the same for all firms in the differentiated sector, i.e. $X_i = X$, $L_i = L$, and $\Pi_i = \Pi$ for all $i = 1, \cdots, N$. The symmetry property allows us to suppress the i-index from here on.

Before characterizing the model developed in this section, we must tie up some loose ends. First, the representative household inelastically supplies H units of labour and is the owner of all firms and thus receives all profits (if there are any). Household income is thus given by:

$$I = HW + N\Pi. \tag{14}$$

The second loose end concerns the labour market clearing condition, according to which the demand for labour by the two sectors must equal the exogenously given supply:

$$NL + L_Z = H. \tag{15}$$

Due to its simple structure, the model can be solved in closed form. We start by noting that equations (12) and (13) can be combined to obtain a simple expression for profit per active firm in the monopolistic sector:

$$\Pi = W\left[(\mu - 1)\, k_Y X - F\right]. \tag{16}$$

---

curves. Dixon and Lawler (1996, p. 223) propose the following technology which features a U-shaped average cost curve:

$$X_i = \begin{cases} 0 & \text{if } L_i \leq F \\ (1/k_Y)\left[L_i - F\right]^{\gamma} & \text{if } L_i \geq F \end{cases}$$

with $0 < \gamma < 1$.

With free entry/exit of firms, profits are driven down to zero and the unique output level per active firm follows directly from equation (16):

$$X = \frac{F}{(\mu - 1)\, k_Y}. \tag{17}$$

Output per firm is constant and depends only on features of the technology (F and k_Y) and on the gross markup ($\mu \equiv \sigma/(\sigma - 1)$). The lower is σ, the higher is μ and the smaller is each firm's output. In terms of Figure 1, the Chamberlinian equilibrium is represented by point E: P^m is given by equation (13) and X^m corresponds to equation (17).

Since profits are zero in the Chamberlinian equilibrium, it follows from (14) that $I = HW$ and from (4) that $Z = \delta HW/P_Z$. By using this result in (9) and (10) we find the equilibrium levels of output and employment in the homogeneous goods sector:

$$Z = \frac{L_Z}{k_Z} = \frac{\delta H}{k_Z}. \tag{18}$$

A constant share of the labour force is employed in the homogeneous goods sector.

From (11) and (17) we find that in the symmetric equilibrium $L = k_Y X + F = \mu k_Y X$ or in aggregate terms $NL = \mu k_Y NX$. By using (15) and (18) we find that $NL = (1 - \delta) H$. Since output per firm is known, we can combine these two expressions for NL and solve for the equilibrium number of firms:

$$N = \frac{(1 - \delta) H}{\sigma F}, \tag{19}$$

where we have used the fact that $\mu \equiv \sigma/(\sigma - 1)$ to simplify the expression somewhat. The equilibrium number of firms depends positively on the amount of labour attracted into the monopolistically competitive sector and negatively on the demand elasticity and the level of fixed cost that each firm must incur. All these effects are intuitive.

Aggregate output of the monopolistically competitive sector can be computed as follows. Equation (2) implies that in the symmetric equilibrium $Y = N^{\mu} X$. By using this result and noting (17) and (19) we find:

$$Y = \Omega_0 L_Y^{\mu}, \tag{20}$$

where $\Omega_0 \equiv (\sigma - 1)\, \sigma^{-\mu} F^{1-\mu}/k_Y$ is a positive constant and $L_Y \equiv (1 - \delta) H$ is the total labour force employed in the monopolistically competitive sector. The key thing to note about equation (20) is that, since $\mu > 1$, labour features increasing returns to scale in the Chamberlinian model. Inspection of (17) and (19) reveals that a larger market (prompted, say, by an increase in the labour force H) leaves the equilibrium firm size unchanged but expands the number of product varieties. Note that by using (7) in the symmetric equilibrium, (10), and (13) we find that the relative price of the composite differentiated good can be written as follows:

$$\frac{P_Y}{P_Z} = \left(\frac{\mu k_Y}{k_Z}\right) N^{1-\mu}. \tag{21}$$

This expression provides yet another demonstration of the scale economies that exist in the Chamberlinian model. These scale economies originate from the love-of-variety effect (see also footnote 17). Provided μ exceeds unity, the relative price of the differentiated good falls as the number of product varieties rises.

An attractive feature of the Dixit-Stiglitz model is that it contains the perfectly competitive case as a special case. Indeed, by letting σ approach infinity and, at the same time, letting F go to zero, both sectors in the economy are perfectly competitive. Since $\mu = 1$ in that case, it follows from (16)-(21) that profits are identically equal to zero ($\Pi = 0$), output per firm and the number of firms are undetermined, aggregate output features constant returns to scale, and the relative price only depends on relative productivity (k_Y/k_Z).

4.1.4 Welfare Properties

Does the Chamberlinian market equilibrium provide too much or too little variety? This is one of the classic questions that has been studied extensively in the monopolistic competition literature. The problem is illustrated by Figure 1. At point E there are unexploited economies of scale due to markup pricing. Salop (1979, p. 152) uses a spatial model of monopolistic competition and concludes that the market produces too much variety. He is careful to note, however, that this result is not robust. In contrast, in the standard Dixit-Stiglitz model the first-best ("unconstrained") social optimum calls for more product varieties than are provided by the market (1977, p. 302)—see also below.[20] Spence reaches the same conclusion in a special case of his model but argues that the problem is inherently difficult to study because:

> ...there are conflicting forces at work in respect to the number or variety of products. Because of setup costs, revenues may fail to cover the costs of a socially desirable product. As a result, some products may be produced at a loss at an optimum. This is a force tending towards too few products. On the other hand, there are forces tending toward too many products. First, because firms hold back output and keep price above marginal cost, they leave more room for entry than would marginal cost pricing. Second, when a firm enters with a new product, it adds its own consumer and producer surplus to the total surplus, but it also cuts into the profits of the existing firms. If the cross elasticities of demand are high, the dominant effect may be the second one. In this case entry does not increase the size of the pie much; it just divides it into more pieces. Thus, in the presence of high cross elasticities of demand, there is a tendency toward too many products. (1976, pp. 230-231)

[20] In the "unconstrained" social optimum, only the resource constraint is taken into account. In the "constrained" social optimum the additional requirement of non-negative profit per active firm is added.

In the remainder of this subsection we study what (our version of) the Dixit-Stiglitz model has to say about this issue.[21]

First-best social optimum In the first-best social optimum the social planner chooses the combination of Z, Y, and N such that the representative household's utility (1) is maximized given the technology (9) and (11) and the resource constraint (15). In the aggregate this problem can be written as:

$$\begin{aligned} \max_{\{Z,Y,N\}} U &= Z^{\delta} Y^{1-\delta} \text{ subject to:} \\ H &= k_Y N^{1/(1-\sigma)} Y + FN + k_Z Z, \\ N &\geq N^{MIN}, \end{aligned} \tag{22}$$

where N^{MIN} is the (exogenous) lower bound on the number of product varieties that can be produced. This lower bound is typically ignored in the literature but is nevertheless important because there are internal scale economies in the differentiated sector. Using the subscript 'u' to denote the first-best optimum values we find the following first-order conditions:

$$\frac{(1-\delta)/Y_u}{\delta/Z_u} = \frac{k_Y N_u^{1/(1-\sigma)}}{k_Z}, \tag{23}$$

$$\left(N_u - N^{MIN}\right)\left[\zeta_u - \frac{\sigma}{\sigma-1}\right] = 0, \tag{24}$$

where ζ_u is a measure for the scale economies at firm level:

$$\zeta_u \equiv \frac{k_Y X_u + F}{k_Y X_u}, \tag{25}$$

and where X_u is production per active firm in the first-best optimum. Equation (23) says that the marginal rate of substitution between the homogeneous and the composite good must be equated to their relative social price. Equation (24) has two important implications. First, if the scale economies at firm level are very strong, and ζ_u exceeds $\sigma/(\sigma-1)$, then the term in square brackets on the left-hand side of (24) is positive and it is socially optimal to produce as few as possible product varieties. In that case the "business-stealing effect"[22] is stronger than the preference-for-diversity effect.

The second implication that can be derived from (24) is that the existence of an internal solution to the optimal number of firms ($N_u > N^{MIN}$) implies that the diversity preference must be sufficiently strong compared to the internal scale economies. Indeed, if the optimal number of varieties is strictly greater than the lower bound, then the term in square brackets

[21] The welfare analysis follows the approach of Broer and Heijdra (2001).

[22] See Mankiw and Whinston (1986, p. 49) for this terminology. The business-stealing effect is operative if output per firm falls when the number of firms increases.

on the left-hand side of (24) is zero (so that $\zeta_u = \sigma/(\sigma-1)$) and (25) determines the optimal size of each active firm:

$$X_u = \frac{F}{k_Y(\mu-1)}, \tag{26}$$

where we have used the definition of μ, given in (13) above, to facilitate the comparison with the corresponding market solution (17). By using the resource constraint and (23) and noting that $Y_u = N_u^{\mu} X_u$ we can solve for the optimal number of firms and the production and employment levels in the homogeneous sector:[23]

$$N_u = \frac{(1-\delta)(\mu-1)}{\delta+\mu(1-\delta)}\frac{H}{F}, \tag{27}$$

$$Z_u = \frac{L_{Yu}}{k_Z} = \frac{\delta}{\delta+\mu(1-\delta)}\frac{H}{k_Z}. \tag{28}$$

Market equilibrium We are now in a position to examine the welfare properties of the model by comparing the first-best socially optimal values of the different variables to their market-produced counterparts (for which we use the subscript 'e'). A comparison of (18) and (28) reveals that $L_{Ze} > L_{Zu}$ and $Z_e > Z_u$, i.e. employment and output in the homogeneous goods sector are too large in the market equilibrium. But this means, of course, that aggregate employment in the differentiated sector is too small, i.e. $N_e L_e < N_u L_u$.

Comparing (17) and (26) shows that $X_e = X_u$, i.e. the market yields the optimal firm size. Finally, by comparing (19) and (27) we find that $N_e < N_u$, i.e. the market provides *too few varieties*! Since output per active firm is the same in the two cases, it follows that $Y_e < Y_u$.[24]

Decentralization It is not difficult to show that the first-best social optimum can be decentralized by means of a product subsidy on differentiated goods, provided it can be financed in a non-distorting fashion (e.g. with a lump-sum tax levied on the representative household). Assume that each active firm receives an ad valorem product subsidy so that equation (12) becomes $\Pi_i \equiv (1+s_P) P_i X_i - W[k_Y X_i + F]$ where s_P is the product subsidy. Each firm will now set its price according to the augmented markup rule $P_i = \mu W k_Y/(1+s_P)$. Since the first-best equilibrium calls for marginal cost pricing ($P_i = W k_Y$) it follows that the optimal product subsidy which decentralizes the first-best social optimum is:

$$s_{Pu} = \mu - 1. \tag{29}$$

Provided the product subsidy is set according to (29), no lump-sum payments to active firms are needed to ensure the optimal amount of variety.

[23] The resource constraint simplifies to $H = k_Z Z_u + \mu k_Y N_u X_u$ and (23) can be simplified to $(1-\delta) k_Z Z_u = \delta k_Y N_u X_u$. Since X_u is known, these two expressions can be solved for N_u and Z_u.

[24] Dixit and Stiglitz are careful to note that this conclusion is specific to the symmetric CES case considered here.

Extensions It is not surprising, in view of Spence's remarks quoted above, that the deficient-diversity result can be easily reversed in a suitably modified version of the Dixit-Stiglitz model. In recent years a number of authors have argued that household preference-for-diversity and market power of individual firms are conceptually distinct phenomena.[25] One way to bring this distinction into focus is by generalizing equation (2) to:

$$Y = N^{\eta-\mu}\left[\sum_{i=1}^{N} X_i^{1/\mu}\right]^{\mu}, \tag{30}$$

where η and $\mu \equiv \sigma/(\sigma-1)$ are allowed to be distinct.[26] In this formulation, μ (>1) parameterizes the market power of differentiated goods producers whereas η (≥ 1) captures the preference-for-diversity effect.[27] If $\eta = 1$ then households do not value variety *per se* but firms nevertheless possess market power as μ exceeds unity. In the standard Dixit-Stiglitz model, $\eta = \mu$, and it is impossible to switch off the preference-for-diversity effect without at the same time eliminating the basis for monopolistic competition.

Following the same steps as before, it is easy to show that the market-equilibrium values for X, Z, and N are still as given in, respectively, (17), (18), and (19). The relative price of the composite differentiated good is given by:

$$\frac{P_Y}{P_Z} = \left(\frac{\mu k_Y}{k_Z}\right) N^{1-\eta}, \tag{31}$$

and aggregate output in the differentiated sector is:

$$Y = \Omega_0' L_Y^{\eta}, \tag{32}$$

where $\Omega_0' \equiv \eta^{-\eta}(\eta-1)^{\eta-1} F^{1-\eta}/k_Y$ and $L_Y = (1-\delta)H$. Comparing (20) and (32) we find that it is the preference-for-diversity parameter, η, which regulates whether or not there are increasing returns to labour. Indeed, if $\eta = 1$, then returns to scale are constant despite the fact that there is monopolistic competition in the differentiated sector.

[25] See Bénassy (1996) and Broer and Heijdra (2001).

[26] The true price index associated with (30) is:

$$P_Y \equiv N^{\mu-\eta}\left[\sum_{i=1}^{N} P_i^{1-\sigma}\right]^{1/(1-\sigma)}.$$

The demand for variety i is now given by:

$$\frac{X_i}{Y} = N^{(\eta-\mu)/(\mu-1)}\left(\frac{P_i}{P_Y}\right)^{-\sigma}.$$

[27] Interestingly, preferences such as (30) actually appear and are analyzed in the May 1974 and February 1975 versions of the Dixit-Stiglitz paper which are reprinted here as Chapters 3 and 4, respectively. Unfortunately, they apparently did not survive the refereeing process and were eliminated from the published paper. They do, however, feature in Ethier (1982) though in a slightly different context.

The welfare analysis is also affected by the alternative definition of Y given in (30). Indeed, it is not difficult to show that the first-best interior social optimum calls for:[28]

$$Z_u = \frac{\delta}{\delta + \eta(1-\delta)} \frac{H}{k_Z}, \tag{33}$$

$$X_u = \frac{F}{(\eta - 1) k_Y}, \tag{34}$$

$$N_u = \frac{(1-\delta)(\eta - 1)}{\delta + \eta(1-\delta)} \frac{H}{F}, \tag{35}$$

$$Y_u = N_u^{\eta} X_u. \tag{36}$$

We can once again compare the socially optimal solutions to their market counterparts. Provided η is strictly greater than unity, it follows from (18) and (33) that the homogeneous sector is too large in the market equilibrium, i.e. too little labour enters the differentiated sector. All the other welfare comparisons depend in a critical manner on the magnitude of the preference-for-diversity parameter, η, relative to the markup μ. Indeed, the comparison of (17) and (34) reveals that $X_e \lessgtr X_u$ for $\eta \lessgtr \mu$. In the first-best social optimum, preference for diversity determines firm size in the differentiated sector, whereas in the Chamberlinian model the markup performs this role. If individual firms possess a lot of market power (σ close to unity so that μ is large), then output per variety is kept small and the market mechanism produces firms that are too small from a social point of view.

The comparison of (19) and (35) yields the following conclusion regarding the number of product varieties:

$$N_e \lessgtr N_u \qquad \Longleftrightarrow \qquad [\sigma - (1-\delta)](\eta - 1) \lessgtr 1 \tag{37}$$

Dixit and Stiglitz (1977, p. 302) study the special case for which $\eta = \mu$. In that case, $\sigma = \eta/(\eta - 1)$ and the market must yield too few varieties (see also above). If, however, we allow η and μ to be different, then the conclusion is less clear-cut. If the diversity effect is strong, and $\eta > \mu$, then the market still produces too few varieties. In contrast, if the diversity effect is weak ($\eta \approx 1$), market power is strong ($\sigma \approx 1$), and the differentiated sector is relatively large ($\delta \approx 0$), then it may well be the case that the N_e exceeds N_u and there is excessive product differentiation. In that case active firms in the market are very small (because they possess a lot of market power) and a large part of the labour force is employed in the differentiated sector.

Decentralization of the first-best equilibrium is more complex in the augmented model. In addition to a product subsidy, the policy maker must now be able to directly affect the zero-profit condition by means of a firm-based lump-sum tax or transfer. In particular, if we denote this instrument by T_u and index it with P_Y, then the profit definition (12) is

[28] We limit attention to the interior solution for which the preference for diversity effect is strong enough to dominate the internal scale economies. It is not difficult to show that condition (24) for the augmented model is $\left(N_u - N^{MIN}\right)[\zeta_u - \eta] = 0$.

augmented to $\Pi_i = (1+s_P)\,P_iX_i - W\,[k_YX_i + F] - P_YT_u$. The first-best equilibrium can be decentralized if the product subsidy is set equal to $s_{Pu} = \mu - 1$ (see equation (29) above) *and* if the lump-sum tax is set according to:

$$T_u = (\mu - \eta)\,N_u^{\eta-1}X_u. \tag{38}$$

The product subsidy restores marginal cost pricing and the lump-sum instrument is needed to ensure that active firms have the socially optimal size and the optimal number of product varieties is produced. The expression in (38) is intuitive. For example, if $\eta > \mu$ then lump-sum transfers are needed. It is optimal to have many small firms but, in the absence of a lump-sum transfer, such firms are unable to cover their fixed costs despite the fact that they receive a product subsidy. The opposite holds if $\mu > \eta$.

If the policy maker does not possess the lump-sum instrument for firms, the first-best cannot be decentralized and a more restricted welfare criterion must be employed. As Dixit and Stiglitz themselves put it, "it would therefore appear that a more appropriate notion of optimality is a constrained one, where each firm must have nonnegative profits" (1977, p. 300). Clearly, in view of our previous discussion, the non-availability of the lump-sum instrument is irrelevant in case $\eta = \mu$—see equation (38). In the general case, however, the second-best (or "constrained") social optimum is indeed different from the first-best ("unconstrained") social optimum.

In the constrained social optimum the planner must find an optimal trade-off between two tasks, namely removing the monopoly distortion (as parameterized by $\mu - 1$) and producing the optimal number of product varieties. For our version of the Dixit-Stiglitz model, the constrained social optimum can be determined as follows. First we note that for a given value of the product subsidy, s_P, the Chamberlinian market equilibrium yields the following solutions:

$$X_e = \frac{F}{(\mu-1)\,k_Y}, \tag{39}$$

$$Z_e = \frac{L_{Ze}}{k_Z} = \frac{\delta}{1+(1-\delta)\,s_P}\frac{H}{k_Z}, \tag{40}$$

$$N_e = \frac{(1-\delta)\,H}{\sigma F}\left(\frac{1+s_P}{1+(1-\delta)\,s_P}\right), \tag{41}$$

$$Y_e = N_e^{\eta}X_e. \tag{42}$$

In the constrained social optimum, the social planner maximizes household utility, $U \equiv Z_e^{\delta}X_e^{1-\delta}N_e^{\eta(1-\delta)}$, taking into account the relations (39)-(42). It is already clear, from the inspection of (17) and (39), that the planner is unable to affect firm size in the constrained equilibrium, i.e. $X_c = X_e$, where the subscript "c" is used to denote the constrained social optimum. The planner can, however, affect both the number of product varieties and the size

of the homogeneous goods sector. By using, respectively, (40) and (41) we find:

$$\frac{1}{Z_e}\frac{\partial Z_e}{\partial s_P} = -\frac{1-\delta}{1+(1-\delta)\,s_P} < 0, \tag{43}$$

$$\frac{1}{N_e}\frac{\partial N_e}{\partial s_P} = \frac{\delta}{(1+s_P)\,[1+(1-\delta)\,s_P]} > 0. \tag{44}$$

Increasing the product subsidy leads to a shrinkage of the homogeneous goods sector and an expansion of the number of firms in the differentiated sector. By differentiating the utility function with respect to s_P we find after some manipulation:

$$\begin{aligned}\frac{dU}{ds_P} &= U\left[\frac{\delta}{Z_e}\frac{\partial Z_e}{\partial s_P} + \frac{\eta\,(1-\delta)}{N_e}\frac{\partial N_e}{\partial s_P}\right] \\ &= \frac{\delta\,(1-\delta)\,U}{1+(1-\delta)\,s_P}\left[-1+\frac{\eta}{1+s_P}\right].\end{aligned} \tag{45}$$

This expression clearly shows that in the constrained social optimum, the preference-for-diversity effect provides the rationale for a product subsidy. This stands in stark contrast with the unconstrained social optimum, in which the product subsidy aims to restore marginal cost pricing and thus depends on the index for market power, $\mu-1$—see equation (29) above. In the constrained social optimum, the policy maker sets the product subsidy such that $dU/ds_P = 0$, or:

$$s_{Pc} = \eta - 1. \tag{46}$$

By using (46) in (40) and (41) and rearranging somewhat, we find the expressions for Z_c and N_c:

$$Z_c = \frac{L_{Zc}}{k_Z} = \frac{\delta}{\delta+\eta\,(1-\delta)}\frac{H}{k_Z}, \tag{47}$$

$$N_c = \left(\frac{(1-\delta)\,(\mu-1)}{\delta+\eta\,(1-\delta)}\right)\frac{\eta}{\mu}\frac{H}{F}. \tag{48}$$

By using (33), (40) (with $s_P = 0$ imposed), and (47) we find that $Z_u = Z_c < Z_e$, i.e. the homogeneous sector has the same size in the unconstrained and the constrained social optimum but is too large in the Chamberlinian equilibrium. The comparison for the number of product varieties proceeds as follows. First, we note from (44) that N_e is increasing in s_P. If $\eta > 1$ then it follows from (46) that $s_P > 0$. As a result, we find that $N_c > N_e$, i.e. the constrained social optimum also calls for a larger number of product varieties. Second, the comparison of (35) and (48) reveals:

$$\frac{N_c}{N_u} = \frac{\eta\,(\mu-1)}{\mu\,(\eta-1)}. \tag{49}$$

In the standard Dixit-Stiglitz case, $\eta = \mu$ and it follows from (49) that $N_c = N_u$. In the generalized model, we find that $N_c \gtreqless N_u$ for $\eta \gtreqless \mu$.

4.2 Evaluation

We have shown above that the Dixit-Stiglitz model, though rather specific in its assumptions, offers a neat way of dealing with increasing returns to scale at firm level whilst maintaining a well-defined industry equilibrium. In addition we have shown how the model captures the key insights of Chamberlin and is rather flexible.

This is not to say that the Dixit-Stiglitz model has not been criticized in the literature. Though the model was originally intended as a contribution to the product differentiation literature, it has come under serious attack in that literature. Eaton and Lipsey, for example, suggest a number of so-called "awkward facts" about product differentiation which the model fails to accommodate (1989, pp. 725-726, 731). First, in reality each consumer only buys a *small subset* of the available commodities. Second, in reality tastes are revealed to differ among individuals in that "...different consumers purchase different bundles of differentiated commodities and these differences cannot be fully accounted for by differences in their incomes" (Eaton and Lipsey, 1989, p. 726). In the Dixit-Stiglitz model there is a *representative* consumer who buys *all* existing varieties. In the product differentiation literature the aim is therefore to go beyond the representative-agent world of Dixit and Stiglitz and to search for microeconomic foundations of the Chamberlin model.[29]

Though the Dixit-Stiglitz model has failed to impress most students of product differentiation, it has nevertheless attained "workhorse" status in a large number of different fields of economics. Especially the analytical flexibility of the model has proved to be very convenient in the last quarter century or so. In Figure 2 we plot the number of citations to the published version of the Dixit-Stiglitz paper.[30] There has been a steady increase in the total number of citations and even in core journals the paper holds its ground. This book shows where these citations are coming from. The specific formalization of the model, and the fact that it can deal with imperfect markets without getting lost in details of strategic interaction, has led to innovations in various fields of economics, as this volume of articles aims to show.

5 The Book

The structure of this book is as follows. The book consists of six parts, of which two are general and four are more specific and each cover a separate discipline. We selected those disciplines where the Dixit-Stiglitz model has contributed the most, namely international trade theory, geographical economics, macroeconomics, and growth theory.

Each discipline part consists of a survey paper that is followed by a critical appraisal of the

[29] Eaton and Lipsey (1989, pp. 731-734) present a brief (and now somewhat dated) survey of some of this micro-foundations literature. In a recent paper, Weitzman (1994) has demonstrated that a generalized aggregator function like (2) can be interpreted as the reduced form of a spatial model of monopolistic competition on the circle, provided the firm can choose its level of specialisation.

[30] We thank Harry van Dalen of Erasmus University for providing us with this figure.

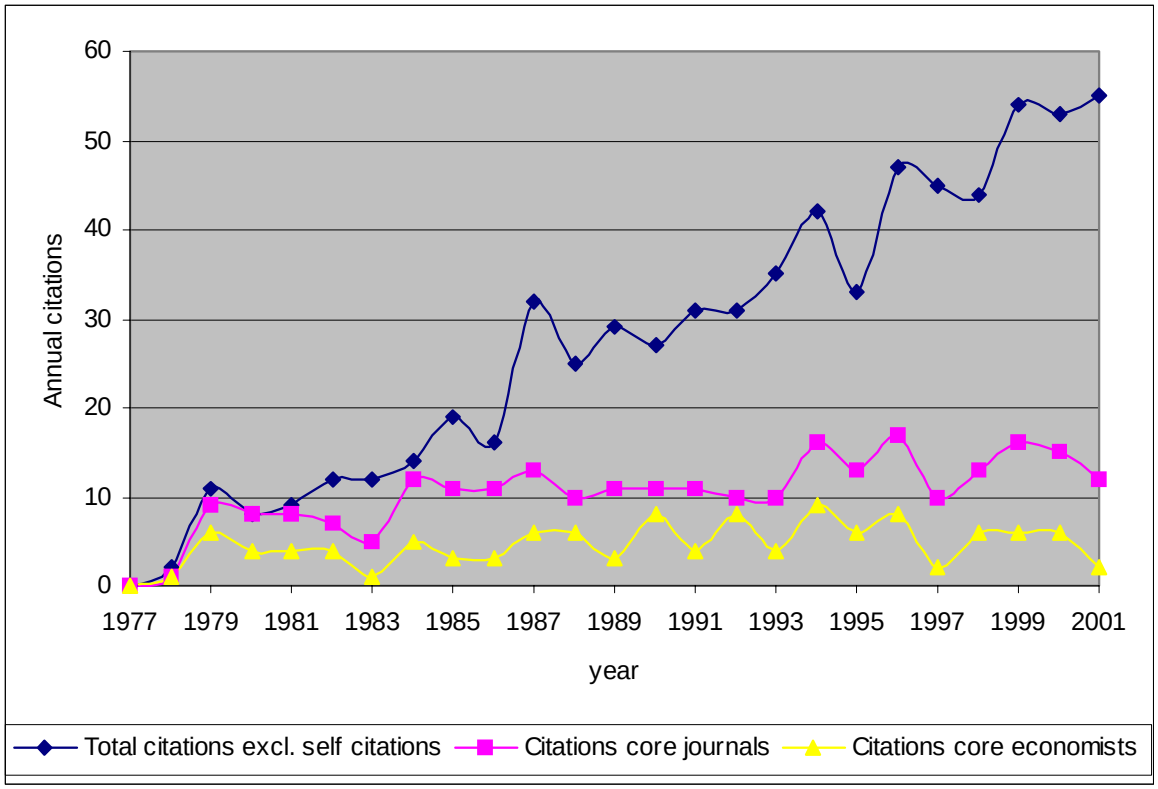

Figure 2: Citations to the Dixit-Stiglitz (1977) paper

Dixit-Stiglitz model in that particular field, and one or two original applications of the Dixit-Stiglitz model. We hope that this volume shows that, although the model has unrealistic features, it nevertheless gives interesting answers to many diverse problems. Furthermore, we hope to show that the second monopolistic competition revolution has indeed been more successful than the first.

Part I: Underground Classics

In Part I we include some "underground classics", i.e. one previously unpublished paper by Stiglitz on the capital market and two preliminary versions of the Dixit-Stiglitz paper dated, respectively, May 1974 and February 1975. These working papers not only develop the basic model in more detail than the published version does, but also contain some interesting material that later had to be, independently, rediscovered by others. For example, the May 1974 version of the model includes the following features: (a) preference for diversity and market power are clearly distinguished (see also the discussion surrounding equation (30) above), and (b) preferences are modelled in terms of a continuum of product varieties. Similarly, the February 1975 version of the paper contains an interesting discussion of the public good aspect of product variety. We wholeheartedly concur with Benassy who argues that especially the February 1975 version of the Dixit-Stiglitz paper "...should become required reading for all serious students of the field" (1996, p. 46f3). By incorporating it (and its May 1974 precursor) in this book we hope to have significantly lowered the barrier to such students.

Part II: Current Perspectives

In Part II of the book, the intellectual founding fathers of the Dixit-Stiglitz model reflect on their contribution to the monopolistic competition literature. In addition, one of the earliest adopters of the Dixit-Stiglitz model, Wilfred Ethier, reflects on the usefulness of this model to international trade theory.

In Chapter 5, Avinash Dixit is rather critical of the Dixit-Stiglitz approach, and wants to go beyond the simplifying assumptions of the benchmark model, such as new perspectives from game theory. He argues that one of the reasons the first monopolistic competition revolution failed was because of the absence of any strategic interaction between firms. In essence this criticism mirrors the ones voiced against Marshall's representative firm almost a century ago. Also in Marshall's analysis strategic interaction is basically absent. Dixit states that this is also true for the second monopolistic competition revolution, where all firms are in essence "price-index" takers, and the determination of this price index is left to higher levels of aggregation. Dixit is in favor of a more eclectic view in which no single model should dominate economic analysis—more than one model is needed to describe a complex reality. What is needed is a taxonomy of different models that can be applied to different situations. And in this sense Dixit welcomes the—sometimes criticized—trend towards numerical simulations

of theoretical models that "make the theoretical models come alive in a way that complex algebraic comparative static expressions do not."

In Chapter 6, Joe Stiglitz recalls that four decades after Chamberlin had written his book, the theory of monopolistic competition called out for formalization. The only formal model of monopolistic competition available at that time was one in which product differentiation showed up as a series of stores located on a circle: each store having only two neighbours with whom to compete. But this particular formalization seems at odds with the central idea of the monopolistic competition model: it was not realistic to assume the absence of (complex) strategic interactions between these stores. Stiglitz remembers that "we wanted to formulate a bench-mark model." He continues that with hindsight it seems that the benchmark model was taken "too seriously", and shows that other modelling perspectives are potentially more promising, such as dealing with the consequences of imperfect information.[31]

In Chapter 7, Wilfred Ethier calls the Dixit-Stiglitz paper "a truly great tool-paper." Ethier (1982) was a rather early application of the Dixit-Stiglitz framework to international trade, though he himself considers his own paper as yet "another tool-paper"—it has never been among his favorites. But obviously this sentiment is not shared by people working in growth theory where the Ethier-formalization of the Dixit-Stiglitz framework has become very popular. Furthermore (as an aside) Ethier, in his contribution to this volume, gives a funny insight into the refereeing process and his perceptive response to one of Krugman's first papers on New Trade Theory.

Part III: International Trade

Background

Until 1975 international trade theory was dominated by the competitive paradigm embodied in the Heckscher-Ohlin-Samuelson (H-O-S) framework. Things changed quite dramatically following the empirical study of Grubel and Lloyd (1975). They showed the importance of intra-industry trade between countries with similar factor endowments. The H-O-S model, with its explicit reliance on differences in relative factor endowments and homogeneous goods, could not explain this empirical regularity. Initially, this led to many mutually unrelated approaches to deal with intra-industry trade—see the *Handbook of International Economics* chapter by Helpman (1984) for a survey. In the words of Paul Krugman, "...inevitably, given the state of the field at that time, the general impression conveyed in that chapter was of a collection of highly disparate and messy approaches, standing both in contrast and in

[31] Similar critical remarks on the benchmark model can be found elsewhere. Fujita, Krugman and Venables (1999, p.6), for example, mention that the popularity of the Dixit-Stiglitz model is "baffling", given the unrealistic nature of this model (they also mention that their book might have been called "Games You Can Play With CES Functions", indicating, inter alia, the benefits of the Dixit-Stiglitz formalization of monopolistic competition).

opposition to the impressive unity and clarity of the constant-returns, perfect competition trade theory" (1995, p. 1244).

During the late 1970s and early 1980s it became clear that the Dixit-Stiglitz approach provided a simple and elegant way of modelling imperfect competition and increasing returns to scale. Using this approach one could avoid being "messy," and for this reason the Dixit-Stiglitz model quickly became the dominant benchmark model in international trade theory. The first applications were by Krugman (1979, 1980) himself and by Dixit and Norman (1980).[32]

Using the Dixit-Stiglitz framework it is quite easy to explain intra-industry trade. The basic idea of Krugman (1980) can be explained with the aid of our simple version of the Dixit-Stiglitz model (see Section 4.1 above). Assume that there are two countries that are identical in all respects except (possibly) for their labour endowments. In the absence of trade, Chamberlinian monopolistic competition ensures that there are $N = (1-\delta)\,H/(\sigma F)$ domestic product varieties and $N^* = (1-\delta)\,H^*/(\sigma F)$ foreign varieties, where H^* is labour supply in the foreign country. The opening up of trade between the two countries ensures that both domestic and foreign consumers can now choose $N + N^*$ product varieties.[33] Provided domestic and foreign consumers exhibit preference for diversity, the greater choice of variety improves welfare to both. The key thing to note is that this welfare effect does not hinge on differences between factor endowments, i.e. it still holds if $H = H^*$ in our example. This framework is now a standard part of economics curriculum. Although real wages might differ between countries of different sizes in autarky, after trade they become the same as each consumer now has the same choice among varieties—see equation (7). In this sense trade completely substitutes for international factor mobility (as in the H-O-S model).

Ethier (1982) argued that most international trade is not in final products (as in the Krugman (1980) model) but rather in intermediate goods. In terms of our model, Ethier interprets Y in equation (1) as a homogeneous commodity which is produced according to the production function (2), where X_i now represents an intermediate input variety. This formulation implies that, in addition to *internal* economies of scale in the production of each intermediate input variety, there are *external* economies at the final goods stage. The intuition behind this so-called *Ethier-effect* is as follows. An increase in the number of input varieties ensures that Y-producers can use a more "roundabout" production process and thus lower

[32] See Krugman (1995) for a survey. Section 9.3 of Dixit and Norman (1980) is attributed to unpublished material by Norman. During the preparation of this book we received the handwritten version of this earlier work by Norman. His paper is incomplete and the policy consequences of the basic model are absent. The model developed in section 9.3 of Dixit and Norman (1980) is very similar to this version. A PDF file of Norman's unpublished paper can be downloaded from: `http://www.eco.rug.nl/~heijdra/norman76.pdf`. Apparently Norman did not attempt to publish his paper as a separate article once it was included in this well-known textbook.

[33] As Krugman (1980, p. 951) points out, costless product differentiation ensures that no (domestic or foreign) firm will want to produce a variety that already exists.

product costs.[34] In Ethier's model, opening up of trade is beneficial because it causes external effects in the production of final goods in both countries. Interestingly, Ethier's formulation has become quite popular in the macroeconomic and growth applications of the Dixit-Stiglitz model—see below.

The book

In Chapter 8, Peter Neary presents a critical survey of monopolistic competition and international trade theory. He discusses trade theory but also gives a clear account of why the Dixit-Stiglitz (1977) model has become so popular: it is its *formalization* that has contributed to its success. Although the model has become standard in trade theory, it has rather mixed success with respect to empirical testing. But not only the limited empirical success of the model makes Neary critical of the whole approach. Like Kaldor before him, Neary argues that "in its assumptions about entry and strategies, monopolistic competition resembles perfect competition much more than it resembles the market structure of many industrial sectors in the real world..." What is needed according to Neary is a GOLE: a theory of General Oligopolistic Equilibrium.[35]

In Chapter 9, Joe Francois and Douglas Nelson present a state-of-the-art overview of all classes of trade models; classical models, the new trade models, and models based on the Ethier framework. They show the differences and similarities between these models and provide a generic representation of the new trade models and Ethier type of models, combining them with neoclassical elements. Their geometric representation of five classes of models shows why some models give identical results and why some do not. This is a complex issue because the Ethier-type models are usually characterized by multiple-equilibria, and instability. Although various authors express the need for additions, extensions, and better empirical work in this area, the Francois-Nelson taxonomy of models and their geometric representation, show that still not all applications of the basic Dixit-Stiglitz framework have been exploited in international trade theory.

Part IV: Geographical Economics

Background

As we already argued above, for otherwise identical countries, a difference between labour force sizes ($H \neq H^*$) implies that the larger country will have a higher real wage in autarky due to the larger number of available varieties (see also equation (7) and footnote 17). In the presence of impedements to trade (such as transport cost or tariffs), however, it is no longer the case that trade substitutes for factor mobility—see Krugman (1979, 1980). The larger region

[34] In fact, Ethier uses a production function like (30) in which η captures the Ethier effect and μ represents the market power by input variety producers.

[35] Recently, Neary (2002) has proposed such a model.

now offers a higher real wage because of a larger *local* market (which does not face transport costs or tariffs). The effect of this seemingly small addition to the model is surprisingly large and leads to a so-called *home-market effect*: the country with the largest economy tends to be a net exporter of products of the monopolistically competitive industry. Helpman and Krugman (1985) show that if one country is large enough compared to its trading partner, all Y (see equation (1)) might be concentrated in that country. In standard trade models location does not matter, because when transport cost are absent there is nothing to gain from being in a specific location. The home-market effect is interesting by itself, but takes the distribution of labor (or income) as given. If workers would be mobile all workers would end-up in the largest market in order to forego transport cost, and the multicountry model would simply collaps into a single country/region model.

A decade ago, Krugman (1991a, 1991b) realized that his trade model, appended with transport costs, could explain center-periphery patterns in the world economy once factor mobility between countries is allowed. According to Krugman (1999),

> "it is obvious-in retrospect-that something special happens when factor mobility interacts with increasing returns (...). This observation is, as I suggested, obvious in retrospect; but it certainly took me a while to see it. Why exactly I spent a decade between showing how the interaction of transport costs and increasing returns at the level of the plant could lead to the "home market effect" (Krugman 1980) and realizing that the techniques developed there led naturally to simple models of regional divergence (Krugman 1991) remains a mystery to me. The only good news was that nobody else picked up that $100 bill lying on the sidewalk in the interim."

In order to prevent the model from collapsing into a single country/region model, Krugman (1991) makes some additional assumptions regarding the labor market. First, he divides (exogenously) the labor market into two parts: a share is allocated to the homogenous good sector Z (agriculture in his terminology), and a share is allocated to Y (which he calls manufactures). To simplify notation Krugman assumes that a share δ goes to Z, and $1-\delta$ to the Y sector (compare with equation (18)). Furthermore, workers in the manufacturing industry are assumed to be mobile between countries, whereas farmers are immobile; and farmers cannot become workers or vice versa.

This implies that we can no longer assume a single wage rate for the homogenous sector and the manufacturing sector (compare with section 4.1.2). Assuming costless trade in agriculture (homogenous sector), prices and wages in agriculture can be used as a numeraire (by convenient choice of units). In principle the nominal wage rate in the Y sector can easily be calculated by asking at what price the optimal amount per firm (as indicated by equation (17)) is exactly met by total demand. Total demand comes from two sources: domestic sales and foreign sales. For consumers the price difference between these two sources consists

of transport costs (and we have to adapt equations (6) and (7) accordingly). This line of reasoning gives us the equilibrium price at which demand equals optimal supply. From this equilibrium price one can derive the corresponding nominal wage rate in the manufacturing sector by using equation (13). The real wage follows from dividing the nominal wage by the exact price index (raised to the power $1-\delta$).

If the real wage between countries differ this will lead to labor migration untill real wages in the manufacturing sector are equalized. Depending on the model parameters, all workers might concentrate in a single region, or might be divided over different regions. By assumption complete agglomeration is not possible, because of the immobile farmers.

Once the \$100 bill had been picked up by Krugman, geographical economics turned into one of the most active fields in international trade theory, although not everyone is convinced of the 'revolutionary' character of this new field.[36]

The book

In Chapter 10, Richard Baldwin, Rikard Forslid, Philippe Martin, Gianmarco Ottaviano and Frédéric Robert-Nicoud develop and explain the key features of the core-periphery model (as the Krugman model is now called). They show that the core-periphery model is much more complex than the standard Dixit-Stiglitz trade model. They furthermore explain some of the more difficult issues in these types of models, such as normalizations, characteristics of equilibria (e.g. their (in)stability), and how the model deals with expectations. The paper shows how two seemingly innocuous changes to the standard Dixit-Stiglitz trade model—the introduction of transport cost and international factor mobility—can give rise to complicated and interesting models.

The debate of the effects of globalization on labour markets often takes place in the H-O-S framework. In Chapter 11, Jolanda Peeters and Harry Garretsen show that the neoclassical framework might not be the most appropriate one to study the consequences of globalization on labour markets in the OECD countries. By allowing for wage rigidities in a geographical economics model, they are able to show that the effects of globalization can be very different in different countries. It is by no means certain that low-skilled workers are always worse off following increased globalization. If globalization can be represented by a reduction in transportation costs, then agglomerating and spreading forces both determine long-run outcomes of the model. This makes the conclusions less clear-cut than those obtained from neoclassical models. Potentially, however, the Peeters-Garretsen model can accommodate the rather different institutional circumstances existing in apparently similar countries. Obviously more empirical research is needed to find out whether their model is indeed better equipped to deal with all kinds of different circumstances.

[36] See Neary (2001) for a critique. Brakman, Garretsen and Van Marrewijk (2001) present an extensive description of the intellectual history of this approach.

In his critical survey of the geographical economics approach, Peter Neary (2001) notes that empirical work in the field is largely absent. Indeed, with a few exceptions, it is almost totally lacking. This might come as a surprise as the great interest in the Dixit-Stiglitz trade model was caused by the existence of a recalcitrant empirical irregularity, namely the phenomenon of intra-industry trade. This phenomenon made researchers look for theoretical models that could explain these facts. In Chapter 12, Steven Brakman, Harry Garretsen, Charles van Marrewijk, and Marc Schramm explain why it is difficult to test geographical economics models. Basically, the problem is that many observations are consistent with many competing models. In general, testing is restricted to some of the implications of the geographical economics models, but often these characteristics are consistent with more than one model. It is however, possible to estimate the structural equations of a variant of the standard model, as has been done by Hanson (1998) for the USA. Brakman et al. test an extended version of the Hanson model for Germany. They show that it is possible to find some empirical support for this model. It is, however, not always clear how to interpret the results due to the possibility of multiple equilibria. Also here more empirical work needs to be done before one can conclude that the geographical economics approach is better able to explain some of the geographical stylized facts than alternative models are.

To conclude this part of the book, in Chapter 13 Vernon Henderson gives a critical appraisal of the geographical economics approach. As a researcher of a neighbouring discipline, urban economics, he discusses the merits of the model with respect to the insights it offers in its urban applications. Henderson welcomes the tendency of mainstream economists to put more geography into their models, but notes that all elements of the Krugman model are familiar to those working in regional economics.[37] Furthermore, modern models in urban economics have incorporated some of the more relevant stylized facts of cities and city formation, whereas the Krugman model only deals with a simplified model of a city—see Fujita and Thisse (2002). The Krugman model, for example, in its basic formulation predicts that the cost of living in a city falls with city size, a prediction that is at odds with the congested reality of cities. Furthermore, the absence of a land development market is also troublesome for urban economists. Also, from a theoretical perspective the geographical economics models, when applied to the analysis of cities, quickly become complex and predict many possible equilibria. And because we have no idea about the size of relevant parameter values, it is not possible to choose between the different equilibria. As a consequence Henderson is very critical of the application of the Dixit-Stiglitz type of geography models to the analysis of cities and highlights some of the, in his eyes, more promising theoretical developments in urban economics.

[37] This point is acknowledged by Krugman himself, who argues that "in fairness it should be reported that many geographers feel that the new literature is only telling them what they already knew, with a few technical gimmicks" (1995, p. 1265).

Part V: Economic Growth

Background

Following a period of relative neglect, the theory of economic growth was given a new impetus in the mid 1980s and early 1990s by authors such as Paul Romer (1987, 1990), Robert Lucas (1988), and Gene Grossman and Elhanan Helpman (1991). The body of literature that has emerged subsequent to—and as a result of—their pioneering efforts is often referred to as the "endogenous growth" literature. Whereas the long-run growth rate is essentially exogenous in the older growth literature in the Solow-Swan tradition, the distinguishing feature of the new literature is that long-run growth is endogenously determined within the model. Whereas the accumulation of human capital is the central feature in the Lucas (1988) model, the Romer-Grossman-Helpman type models all make explicit use of the Dixit-Stiglitz approach in one way or another.

We can use our version of the Dixit-Stiglitz model to illustrate the way in which endogenous growth emerges in (a simplified version of) the model by Grossman and Helpman (1991, ch. 3).[38] In this model R&D activities form the basis of economic growth. There is a single homogeneous good, Y, which is produced with the production function (2), where X_i is a differentiated input variety (as in the Ethier approach). The number of existing input varieties, N, is predetermined at any moment in time but can be increased by means of labour-consuming R&D activities. Labour is the only production factor and ownership titles on patents present the only financial asset for savings purposes. Intermediate input producers each hold a blueprint telling them how to produce their own, slightly unique, variety X_i and act as Dixit-Stiglitz monopolistic competitors using the technology $X_i = L_i/k_X$, where L_i is labour use and k_X is a technology index.[39] Technology in the perfectly competitive R&D sector is given by:

$$\dot{N} = \frac{NL_R}{k_R}, \tag{50}$$

where $\dot{N} \equiv dN/dt$ is the number of *new* blueprints (patents), k_R is a productivity index, and L_R is the amount of labour used in the R&D sector. Equation (50) constitutes the "engine of growth" in the model. It incorporates the assumption, due to Romer (1990), that the stock of existing blueprints positively affects the productivity of researchers. As Romer puts it, "[t]he engineer working today is more productive because he or she can take advantage of all the additional knowledge accumulated as design problems were solved during the last 100 years" (1990, pp. S83-84).

The infinitely-lived representative household chooses its optimal time profile for consumption in order to maximize the lifetime utility function, $\Lambda \equiv \int_0^\infty \log C(\tau)\, e^{-\rho\tau} d\tau$, subject to

[38] This example is taken from Benassy (1998). See also Heijdra and van der Ploeg (2002, pp. 461-466) for details of the solution method.

[39] There are no fixed costs.

the budget identity $C + P_N\dot{N} = WH + N\Pi$, where C is consumption of the homogeneous good (equalling Y in equilibrium), ρ is the rate of time preference, P_N is the market value of a patent, W is the real wage rate, and Π is profit per intermediate input firm. The budget identity shows that income (right-hand side) is spent on the consumption of final goods and on the accumulation of patents (i.e. on saving). The optimal savings plans of the household in effect determines the fraction of labour allocated to the R&D sector and thus, via equation (50), the rate of innovation in the economy. It can be shown that the endogenous growth rates for the number of input varieties and for aggregate output are:[40]

$$\frac{\dot{N}}{N} = \frac{(\mu - 1)\, H}{\mu k_R} - \frac{\rho}{\mu}, \qquad \frac{\dot{Y}}{Y} = (\mu - 1)\, \frac{\dot{N}}{N}. \tag{51}$$

The innovation rate increases with the monopoly markup and the side of the labour force but decreases with the rate of time preference. Intuitively, the markup-effect operates as follows. If the markup is high then input producers keep their output small so that, for a given number of varieties, the amount of labour absorbed by the input-producing sector is small.[41] As a result, the R&D sector is relatively large so that the economy innovates at a high rate.

The example presented here, of course, represents no more than the tip of the iceberg. It nevertheless serves to illustrate yet another area where the Dixit-Stiglitz approach has proven to be very useful. Indeed, the mere analytical flexibility of the Dixit-Stiglitz model has enables growth theorists to incorporate the notion of monopolistic competition within a fully specified dynamic general equilibrium model with optimizing households and firms.

The book

In Chapter 14, Sjak Smulders and Theo van de Klundert confirm that the Dixit-Stiglitz model of monopolistic competition has been an important building block for a number of theories of economic growth. Their paper examines the role played by monopolistic competition in several theories of economic growth. They discuss these theories in a general equilibrium framework with two types of research and development (R&D). First, new product lines can be introduced by incurring a sunk cost. Second, incumbent firms can raise productivity by in-house investment in tacit knowledge. Special cases of the model include a dynamic version of the Dixit-Stiglitz model, the semi-endogenous growth model, the semi-endogenous growth model based on variety expansion, the endogenous growth model based on in-house R&D, and a combination of the latter two. It is shown that the intensity of competition play quite a different role in the various cases distinguished.

[40] It is assumed that not all labour is absorbed by the monopolistically competitive sector, i.e. some labour is employed in the R&D sector.

[41] It can be shown that there is no transitional dynamics in $L_X \equiv k_X NX$ and that $L_X = (H + \rho k_R)/\mu$. The results mentioned in the text follow from this expression. Note also that output per intermediate firm declines according to $\dot{X}/X = -\dot{N}/N$, i.e. it get smaller and smaller as time goes on.

In Chapter 15, Sjak Smulders studies the effects of capital mobility on welfare and the speed of adjustment in a two-country growth model. Research and development (R&D) allows monopolistic firms to improve their productivity level whilst national and international knowledge spillovers affect the returns to R&D. The two countries considered differ only with respect to the initial productivity level. The country with the lowest productivity level gradually catches up with the leading country. There is complete convergence in the long run if there is no capital mobility. Under perfect capital mobility, countries end up with equal long-run productivity levels, but permanent differences in consumption. The speed of convergence is larger with perfect capital mobility than with balanced trade. The difference increases with substitution between product varieties and the rate of intertemporal substitution. Capital mobility harms (benefits) the leader (lagging) country if domestic spillovers are more important than international spillovers.

In Chapter 16, Henri de Groot, Marjan Hofkes, and Peter Mulder apply the Dixit-Stiglitz approach in yet another area. Their paper studies the adoption and diffusion of energy-efficient technologies in a vintage model. An important characteristic is that vintages are complementary; there are return to diversity of using different vintages. De Groot et al. argue that this is a potentially relevant part of the explanation of the energy-efficiency paradox. They also analyse how diffusion patterns and adoption behaviour are affected by learning-by-doing and tax policies. It is shown that the stronger the complementarity between different vintages and the stronger the learning by doing, the longer it takes before firms scrap (seemingly) inferior technologies.

Part VI: Macroeconomics

Background

Perhaps the two main dividing issues among macroeconomists are (1) the assumed coordination power of markets and (2) the degree of flexibility of prices and wages. Economists who derive their inspiration from the classicals tend to exhibit a great belief in frictionless markets and emphasize price and wage flexibility. At the other end of the spectrum, economists working in the Keynesian tradition tend to display a higher tolerance for imperfect markets and emphasize that prices and wages may be sticky.

During the latter half of the 1980s, economists from the Keynesian group applied the Dixit-Stiglitz model to study some classic macroeconomic questions. For example, Akerlof and Yellen (1985a, 1985b) and Blanchard and Kiyotaki (1987) develop models with monopolistically competitive price and/or wage setters. They show that price and/or wage stickiness may (a) be optimal to individual price/wage setters and (b) be a general equilibrium phenomenon if there exist small (non-convex) costs associated with price/wage changes, so-called *menu costs*. Intuitively, since under monopolistic competition objective functions are flat at the top, such menu costs ensure that quantity adjustment replaces price adjustment following

an exogenous shock. Since the economy is not in a first-best equilibrium to start with, there are first-order welfare effects associated with second-order-small menu costs.

The monopolistic competition model has also been used in a setting of flexible prices/wages. Kiyotaki (1988), for example, studies the 'vintage-Keynesian' notion of animal spirits and multiple, Pareto-rankable, equilibria. He builds a two-period model incorporating monopolistic competition in the goods market and real investment decisions by entrepreneurs and shows that there exist two rational expectations equilibria, namely a 'pessimistic' one and an 'optimistic' one. He thus shows that the state of entrepreneurs' expectations regarding the future can have important implications for allocation and welfare in a monopolistically competitive world.

The monopolistic competition model has also been used to provide foundations for the 'Keynesian' multiplier—see Mankiw (1988) and Startz (1989).[42] In all macroeconomic applications of the Dixit-Stiglitz approach, the endogenous labour supply response plays a vital role. For the multiplier model this can be illustrated quite easily with the aid of the simple model presented in Section 4.1 above.[43] Instead of choosing two types of goods (as in equation (1) above), the representative household now chooses the composite differentiated good, Y, and leisure, $1-H$ (where 1 is the time endowment and H is labour supply). The household has a Cobb-Douglas utility function, $U=(1-H)^{\delta}Y^{1-\delta}$. The household budget constraint (3) is augmented to $P_Y Y+W(1-H)=I$, where $I\equiv W+N\Pi-T$ is full income, W is the wage rate, $N\Pi$ is aggregate profit income, and T is the lump-sum tax. Equations (5)-(6) are still relevant but (4) is replaced by $W(1-H)=\delta I$. The government consumes a composite good G (defined analogously to (2)) and its budget constraint will be $T=P_Y G$, where P_Y is defined in (7). Individual firms face a demand from both the government and the representative household but since these demand curves feature the same elasticity the price is the same for both customers, i.e. $P_i=\mu W k_Y$ (see (13) above). In the symmetric equilibrium, the goods market equilibrium condition is $NPX=P_Y[Y+G]$ and aggregate profit income can be written as $N\Pi=\frac{1}{\sigma}NPX-WNF$.

We follow Mankiw (1988) by assuming that the number of firms is constant in the (very) short run. In that case, the model can be condensed to two simultaneous equations determining composite household consumption, Y, and real aggregate output, NPX/P_Y, as a function of the exogenous variables:

$$\frac{NPX}{P_Y} = Y+G, \tag{52}$$

$$Y = \Omega_0+\left(\frac{1-\delta}{\sigma}\right)\frac{NPX}{P_Y}-(1-\delta)G, \tag{53}$$

[42] Dixon (1987) remarks that this multiplier is more Walrasian than Keynesian because perfect flexibility of prices and wages is assumed.

[43] The model is loosely based on Mankiw (1988). Details of the solution approach can be found in Heijdra and van der Ploeg (2002, pp. 359-377).

where $\Omega_0 \equiv (1-\delta) W (1-NF) / P_Y$ is a constant. The system is reminiscent of the Samuelsonian Keynesian Cross diagram—equation (52) says that aggregate output equals demand by the household and the government whilst (53) shows that household demand itself depends positively, via its effect on aggregate profit income, on real aggregate output. It is therefore not surprising that the model features a short-run multiplier effect. Indeed, it follows from (52)-(53) that the short-run output multiplier is given by:

$$0 < \frac{d\left(\frac{NPX}{P_Y}\right)}{dG} = \frac{\delta}{1-\frac{1-\delta}{\sigma}} < 1. \tag{54}$$

The intuition behind the multiplier is as follows. The increase in government consumption necessitates an increase in the lump-sum tax which renders households poorer. Since Y and $1-H$ are both normal goods, the consumption of both is reduced. The reduction in leisure consumption is effectuated by an increase in labour supply. This enables the expansion of aggregate output. Note that under perfect competition (with $\sigma \to \infty$) fiscal policy would also increase aggregate output though by less. The key mechanism under both perfect and monopolistic competition is the labour supply response by households.

The book

In Chapter 17, Russell Cooper reviews the contribution of monopolistic competition to macroeconomics. He begins by assessing the various types of theoretical structures that admit monopolistic competition in either product or factor markets. He then studies the quantitative implications of this form of market power as well as implications for the conduct of fiscal and monetary policy. Cooper concludes that macro models based on monopolistic competition are useful because they provide a source of inefficiency (that is not present under perfect competition) and can be used to study price setting behaviour by firms. From a quantitative perspective, however, the fact that markets are monopolistically competitive—rather than perfectly competitive—does not seem to matter very much.

In Chapter 18, Jan Boone gives a formalization of defensive and enterprising strategies for firms. He asks the question whether a rise in competition tends to make firms more enterprising or more defensive. Distinguishing three ways in which competition can be intensified, Boone finds the following results. First, a rise in the number of opponents and a reduction in opponents' costs both make a firm more defensive. Second, an increase in the aggressiveness of interaction between firms makes the top firms in an industry more enterprising while the laggards become more defensive. These results are evaluated in light of the discussion on green production and downsizing.

In Chapter 19, Christian Keuschnigg presents an intertemporal equilibrium model of monopolistic competition and start-up investment with variable capacity. Reflecting a trade-off between the number and capacity of new machines, aggregate investment may be extensive or

intensive and may therefore advance the degree of either specialization or rationalization of industrial production. Investment externalities are identified that result in under-accumulation of capital. The paper compares the effectiveness of a proportional investment tax credit with a fixed start-up subsidy that shifts the direction of investment towards a more extensive form with a larger number of smaller machines.

Finally, in Chapter 20 Leon Bettendorf and Ben Heijdra construct a dynamic overlapping-generations model of a semi-small open economy with monopolistic competition in the goods market. Using the Dixit-Stiglitz framework, they show that there are two distortions that must be addressed by the policy maker, namely the one due to increasing returns to scale (resulting from monopolistic competition) and the one due to national market power (resulting from a downward sloping export demand function). It is natural in this setting to consider two policy instruments, namely the product subsidy and the import tariff. While in the first-best situation the instrument targeting principle is relevant, it turns out that in the second-best case the two policy instruments are complementary. The Dixit-Stiglitz framework thus yields precise and intuitively understandable prescriptions about the interaction between the optimum tariff and pre-existing uncorrected domestic distortions and *vice versa*.

References

Akerlof, G. A. (2002). 'Behavioral macroeconomics and macroeconomic behavior', *American Economic Review*, 92: 411-433.

Akerlof, G. A. and J. Yellen (1985a). Can small deviations from rationality make significant differences to economic equilibria?, *American Economic Review*, 75: 708-721.

Akerlof, G. A. and J. Yellen (1985b). A near-rational model of the business cycle, with wage and price inertia, *Quarterly Journal of Economics, Supplement*, 100: 823-838.

Archibald, G. C. (1961). 'Chamberlin versus Chicago', *Review of Economic Studies*, 24: 2-28.

Archibald, G.C. (1987). Monopolistic competition, in J. Eatwell, M. Milgate, and P. Newman, editors, *The New Palgrave: A Dictionary of Economics*, London, Macmillan.

d'Aspremont, C., Dos Santos Ferreira, R., and Gérard-Varet, L.-A. (1996). On the Dixit-Stiglitz model of monopolistic competition. *American Economic Review*, 86:623–629.

Bénassy, J.-P. (1996). Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters*, 52:41–47.

Bénassy, J.-P. (1998). Is there always too little research in endogenous growth with expanding product variety? *European Economic Review*, 42:61–69.

Bishop, R. L. (1967). Monopolistic competition and welfare economics, in Kuenne, R.E., editor, *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*. John Wiley & Sons, New York.

Blaug, M. (1997). *Economic Theory in Retrospect*, 5th edition. Cambridge University Press, Cambridge.

Brakman, S. and C. van Marrewijk (1998). *The Theory of International Transfers*. Cambridge University Press, Cambridge.

Brakman, S., H. Garretsen, and C. van Marrewijk (2001). *An Introduction to Geographical Economics*. Cambridge University Press, Cambridge.

Broer, D. P. and Heijdra, B. J. (2001). The investment tax credit under monopolistic competition. *Oxford Economic Papers*, 53: 318–351.

Chamberlin, E. H (1933). *The Theory of Monopolisic Competition: A Re-orientation of the Theory of Value*, 8th edition (1962). Harvard University Press, Cambridge, MA.

Cournot, A. (1838), *Recherches sur les Principles Máthematique de la Théorie des Richesses*, Paris.

Dixit, A. and V. Norman (1980). *The Theory of International Trade*. Cambridge University Press, Cambridge.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67:297–308.

Dixit, A. K. and Stiglitz, J. E. (1993). Monopolistic competition and optimum product diversity: Reply. *American Economic Review*, 83:302–304.

Dixon, H.D. (1987). A simple model of imperfect competition with Walrasian features, *Oxford Economic Papers*, 39:134-160.

Dixon, H.D. and Lawler, P. (1996). Imperfect competition and the fiscal multiplier. *Scandinavian Journal of Economics*, 98: 219-231.

Eaton, B.C. and R.G. Lipsey (1989). Product differentiation, in R. Schmalensee and R.D. Willig, editors, *Handbook of Industrial Organization*, Vol. 1, North-Holland, Amsterdam.

Ethier, W. (1982). National and international returns to scale in the theory of international trade, *American Economic Review*, 72: 389-405.

Fellner, W. (1967). The adaptability and lasting significance of the Chamberlin contribution, in Kuenne, R.E. editor, *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Friedman, M. (1953). The methodology of positive economics, in *Essays in Positive Economics*, Chicago, Chicago University Press.

Fujita, M., P. Krugman, and A. J. Venables (1999). *The Spatial Economy*, MIT Press, Boston.

Fujita, M. and J-F Thisse (2002). *Economics of Agglomeration; Cities, Industrial Location, and Regional Growth*, Cambridge University Press, Cambridge.

Grossman, G.M. and E. Helpman (1991). *Innovation and Growth in the Global Economy*. MIT Press, Cambridge, MA.

Grubel, H., and P. Lloyd (1975). *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, MacMillan, London.

Hanoch, G. (1975). The elasticity of scale and the shape of average costs, *American Economic Review*, 65: 492-597.

Harrod, R.F. (1967). Increasing returns, in Kuenne, R.E (ed.), *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Hanson, G.H. (1999), Market Potential, Increasing Returns, and Geographic Concentration, NBER Working Paper Nr. 6429, February 1998 (rev. 1999).

Heijdra, B.J. and F. van der Ploeg (2002). *Foundations of Modern Macroeconomics.* Oxford University Press, Oxford.

Helpman, E. (1984). Increasing returns, imperfect markets, and trade theory, in R.Jones and P.Kenen (eds.), *Handbook of International Economics*, Vol. 1, North-Holland, Amsterdam.

Helpman, E. and P. Krugman (1985). *Market Structure and Foreign Trade*, MIT Press, Cambridge, MA.

Henderson, J.M. and R.E. Quandt (1971), *Microeconomic Theory*, Second Edition. New York, McGraw-Hill.

Hicks, J.R. (1939). *Value and Capital.* Oxford, Oxford University Press.

Johnson, H.G. (1967). International trade theory and monopolistic competition theory, in Kuenne, R.E. editor, *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Kaldor, N. (1934). Mrs. Robinson's 'Economics of Imperfect Competition', *Economica*, 1: 335-341.

Kaldor, N. (1935). Market imperfections and excess capacity, *Economica*, 2: 33-50.

Kreps, D.M. (1990). *A Course in Microeconomic Theory,* Princeton University Press, Princeton.

Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade, *Journal of International Economics*, 9: 469-480.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade, *American Economic Review*, 70: 950-959.

Krugman, P. (1991a). *Geography and Trade*, MIT Press, Cambridge, MA.

Krugman, P. (1991b). Increasing returns and economic geography, *Journal of Political Economy*, 99:483-499.

Krugman, P. (1995). Increasing returns, imperfect competition and the positive theory, in G.M.Grossman and K.Rogoff, editors, *Handbook of International Economics*, Vol. 3, North-Holland, Amsterdam.

Krugman, P. (1999). Was it all in Ohlin?, mimeo, MIT, Cambridge. Downloadable from: `http://www.wws.princeton.edu/~pkrugman/ohlin.html`.

Kuenne, R.E. editor (1967), *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Lucas, R.E. (1988). On the mechanics of economic development, *Journal of Monetary Economics*, 22:3-42.

Malinvaud, E. (1972). *Lectures on Microeconomic Theory*, Amsterdam, North-Holland.

Mankiw, N.G. (1988). Imperfect competition and the Keynesian cross, *Economics Letters*, 26: 7-13.

Mankiw, N.G. and M.D. Whinston (1986). Free entry and social inefficiency, *Rand Journal of Economics*, 17: 48-58.

Marshall, A. (1920), *Principles of Economics*, 8th Edition (reprinted edition, 1994), MacMillan, London.

Neary, J.P.(2001). Of hypes and hyperbolas: Introducing the new economic geography, *Journal of Economic Literature*, 39: 536-561.

Neary, J.P. (2002). The road less travelled: Oligopoly and competition policy in general equilibrium. Forthcoming in: R. Arnott, B. Greenwald, R. Kanbur, and B. Nalebuff, editors. *Imperfect Economics: Essays in Honour of Joseph Stiglitz*, MIT Press, Cambridge, MA.

Obstfeld, M., and K. Rogoff (1996). *Foundations of International Macroeconomics*, MIT Press, Cambridge, MA.

Robinson, J. (1933). *The Economics of Imperfect Competition*, MacMillan, London.

Romer, P.M. Growth based on increasing returns due to specialization, *American Economic Review, Papers and Proceedings*, 77:56-62.

Romer, P.M. Endogenous technological change, *Journal of Political Economy*, 98:S71-S101.

Salop, S.C. (1979), Monopolistic competition with outside goods, *Bell Journal of Economics*, 10: 141-156.

Samuelson, P.A. (1967), The monopolistic competition revolution, in Kuenne, R.E. editor (1967), *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Schumpeter, J.A. (1954). *History of Economic Analysis*, Oxford University Press, Oxford.

Scitovsky, T. (1954). Two concepts of external economies, *Journal of Political Economy*, 62: 143-151.

Spence, A.M. (1976). Product selection, fixed costs, and monopolistic competition, *Review of Economic Studies*, 43: 217-235.

Startz, R. (1989). Monopolistic competition as a foundation for Keynesian macroeconomics, *Quarterly Journal of Economics*, 104: 737-752.

Stigler, G.J. (1949). *Five Lectures on Economic Problems*. Longmans, Green, and Co., London.

Tinbergen, J. (1967). Quantitative economics, macroeconomics, and monopolistic competition, in Kuenne, R.E. editor (1967), *Monopolistic Competition Theory: Studies in Impact, essays in honor of Edward H. Chamberlin*, John Wiley & Sons, New York.

Triffin, R. (1940), *Monopolistic Competition and General Equilibrium Theory*, Cambridge, MA: Harvard University Press.

Weitzman, M.L. (1994). Monopolistic competition with endogenous specialization. *Review of Economic Studies*, 61:57-80.

Yang, X. and Heijdra, B. J. (1993). Monopolistic competition and optimum product diversity: Comment. *American Economic Review*, 83:295–301.